

**UNITED STATES PATENT APPLICATION FOR  
SYSTEM AND METHOD FOR MULTI-LANGUAGE EXTENSIBLE  
COMPILER FRAMEWORK**

Inventor:

Kevin Zatloukal

**CERTIFICATE OF MAILING BY "EXPRESS MAIL"  
UNDER 37 C.F.R. § 1.10**

"Express Mail" mailing label number: EV 386447445 US  
Date of Mailing: 2/19/04, 2004

I hereby certify that this correspondence is being deposited with the United States Postal Service, utilizing the "Express Mail Post Office to Addressee" service addressed to Mail Stop PATENT APPLICATION, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 and mailed on the above Date of Mailing with the above "Express Mail" mailing label number.



Tina M. Galdos  
Signature Date: 2/19/04, 2004

SYSTEM AND METHOD FOR MULTI-LANGUAGE EXTENSIBLE  
COMPILER FRAMEWORK

Inventor: Kevin Zatloukal

COPYRIGHT NOTICE

**[0001]** A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure, as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

CLAIM OF PRIORITY

**[0002]** This application claims priority from the following application, which is hereby incorporated by reference in its entirety:

**[0003]** U.S. Provisional Application No. 60/449,991, entitled SYSTEMS AND METHODS FOR A MULTI-LANGUAGE EXTENSIBLE COMPILER FRAMEWORK, by Kevin Zatloukal, filed on February 26, 2003 (Attorney Docket No. BEAS-01396US0).

TECHNICAL FIELD OF THE INVENTION

**[0004]** This invention relates in general to the field of software systems, specifically software systems for compiling computer programs.

BACKGROUND

**[0005]** A compiler is a computer program that takes as input a computer program written a source language and produces as output an equivalent computer program written in a target

language. It may be designed to translate any source language into any target language. Many compilers, however, are designed to accept only one source and one target language. The source and target languages of these compilers are selected when the compiler is first written. Changing them is nearly impossible and would require a rewrite of virtually the entire compiler.

[0006] Recent trends in the computer industry have been towards more complicated computer programs, often written in multiple computer languages. Furthermore, multiple computer languages might appear in a single source file, often with one language nested inside another. Traditional multiple language compilers are not sufficient to deal with this problem. Some of them were designed to deal with multiple languages in a single source file in limited examples, but none of them deal with the problem in a general way. Furthermore, such compilers cannot be easily extended to support new languages or new combinations of languages in a source file.

[0007] The demands on compilers are increasing in other ways as well. In the past a compiler was designed to serve a single client, typically command line interface, to perform batch compilation of a group of files. Modern compilers are facing more diverse clients which require far more detailed information from compiler. These clients include the traditional batch mode user interfaces as well as integrated development environments.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0008] **Figure 1** is a diagram showing a multi-programming-language compiler system that can be used in accordance with one embodiment of the present invention.

#### DETAILED DESCRIPTION

[0009] The invention is illustrated by way of example and not by way of limitation in the figures of the accompanying drawings in which like references indicate similar elements. It

should be noted that references to “an” or “one” embodiment in this disclosure are not necessarily to the same embodiment, and such references mean at least one.

**[0010]** One embodiment of the present invention provides a system and method for creating a compiler system **100** as shown in **Figure 1** that allows the compilation of multiple languages written in a computer program **101** and can be easily extended to add new languages. It consists of a compiler framework **102**, which creates a general environment in which to carry out compilation, a plurality of language modules such as **103** to encapsulate the details of various programming languages, and a plurality of language interfaces such as **106** provided by each language module to interact with the compiler framework. In order to compile a program, the compiler framework and the language modules operate together to execute the compilation process. The compiler framework controls the compilation process and performs a standard language-independent portion of the compilation process and each language module provides a language-dependent portion of the compilation process for a particular programming language. Such a system makes it easy for tool vendors and end users to adapt to a world where computer programs are written in multiple languages. New language modules may be written that add support for new languages to interact with the compiler framework. It may also be possible to extend existing language modules so that a variant on an existing language may be added to the compiling system.

**[0011]** One embodiment of the present invention may be adapted to permit one or more clients **110** to interact with the compiler system through an information interface **109** in order to request services and obtain detailed language information from the compiler framework. These clients may include a standard command-line shell **112** or a sophisticated multi-language integrated development environment (IDE) **111**. Information from the language modules and the compiler framework may be passed through to the various clients in a language-neutral way.

**[0012]** The compiler framework in accordance with one embodiment of the present invention is responsible for performing services that are not highly specific to any one programming language in the computer program. In some embodiments of this invention the compiler framework may be tailored for a particular environment such as the Java environment. In such a circumstance, the compiler framework may provide services that are more useful for Java-like programming language, but it does not mean that the compiler framework will become language-dependent.

**[0013]** In one embodiment, a computer program **101** that is compiled by the compiler system may be organized into projects. A project may include at least one set of files, paths, libraries, configuration information, and dependencies of files. Such information may be maintained and used by the compiler framework to direct the compilation process. In an embodiment in the Java environment a project might include a list of class files, Java files, JAR files, and a set of Java classpaths.

**[0014]** In one embodiment the compiler framework is responsible for controlling the overall compilation process for a computer program. The phases of the compilation process may be defined by the compiler framework and may include scanning, parsing, name resolution, semantic checking, and code generation. The compiler framework may control the invocation of these phases by calling functions on an interface provided by the language modules.

**[0015]** The compiler framework in accordance with one embodiment of the present invention may maintain a type cache to store types defined in the files of the project. This type cache may allow types defined in different languages to be intermixed and may allow types defined in one programming language to reference types defined in another programming language. In an embodiment for the Java environment, this type cache may maintain a hierarchical structure mirroring the package structure of the Java project. The type cache may also requires types defined in different programming languages to be mapped to the type system

of one particular programming language, such as the Java programming language. In one embodiment, the type cache may contain all the public information about a particular source file so that another source file may be type checked using only the information contained in the type cache.

[0016] A type cache may also store dependencies between the types it stores. A dependency represents the fact that one type may depend in some way on the structure of another type. The compiler framework may also maintain a list of reverse dependencies, making it efficient to determine what other types may be affected if a particular type is changed. The type cache may be serialized to disk so that it does not have to be regenerated when the compiler framework is shut down and restarted.

[0017] The compiler framework may also include a list of errors. In embodiments that organize computer programs into projects and files, the set of errors may include errors for the entire project and errors for each source file in the project. The errors may also include one or more suggestions for correcting the errors, which may be provided by the language modules or the compiler framework.

[0018] The compiler framework in accordance with one embodiment of the present invention may also provide a multi-threading service which may be used by the compiler framework and the language modules. The multi-threading service may include a thread pool with multiple worker threads capable of being assigned to independent tasks. The multi-threading service may also include a facility for including dependencies between the worker threads so that one worker thread may wait on the completion of a second worker thread.

[0019] While the compiler framework is programming language independent, it may be tailored for a particular programming language environment. In one embodiment, the compiler framework may be tailored to the Java programming environment. In this embodiment, the compiler framework may organize use a project system that includes the Java package structure.

The compiler framework may also utilize a Java-like type system for all its programming languages. The framework may also provide a module for code generation that uses Java as an intermediate language.

**[0020]** In one embodiment, the compiler framework may interact with a particular language module through a standard language interface that every language module must implement. This interface might provide functions allowing the compiler framework to access various components that perform different phases of compilation and it may also allow the compiler framework to get language specific information about the source files that have been compiled.

**[0021]** In one embodiment, the language interface may present the language-dependent portion of the compilation process in the form of a set of components, each component performing one of the standard phases of compilation. These phases may include a scanning phase, a parsing phase, a name resolution phase, a semantic checking phase, and a code generation phase.

**[0022]** In one embodiment, the language interface allows one language module to interact with another language module to provide services for compilation of nested languages. Language nesting occurs when a section of source code written in an inner language appears within the source code of an outer language. One family of nested languages consists of the Java annotation languages, where Java is the outer language and the inner language appears within Java comments. The language interface allows one language module to invoke another language module in order to compile a nested language. The outer language may identify the start of a nested language using any information generated during compilation or it may allow the inner language to make the determination. Either the inner or the outer language may determine where the nested language ends.

**[0023]** In one embodiment, the language interface may include functions for retrieving information about a particular source file. These interfaces may provide various types of information that can be used by various clients of the compiler framework. In an embodiment where an integrated development environment (IDE) is a client of the compiler framework, this information may be useful for providing various editing features for the language. In such an embodiment this information may include: information about matching tokens, the list of tokens for a particular source file or a particular portion of a source file, code completion information, or language nesting information.

**[0024]** A language module is the mechanism by which the compiler framework is extended. A language module should encapsulate the knowledge about a particular programming language and present a standard language interface to the compiler framework. A language module controls the portions of the compilation process that require specific knowledge of a programming language. Language modules may be provided by the developer of the compiler framework, by independent vendors, or by an end user.

**[0025]** In one embodiment, one of the language modules might be a language module for the Java language. This Java language module would include several components which have specific knowledge of the Java language. These components might include: a scanner, a parser, a name resolver, a semantic checker, and a code generator each of which has a detailed understanding of part of the structure of the Java language. These components would be invoked by the compiler framework in the necessary order to perform compilation of a Java file.

**[0026]** In one embodiment, one language module may be able to extend another language module in order to easily create a new programming language. For instance, a language like Java could be given extra semantics that are not present in the original language. In embodiments where the language modules provide separate components for each phase of compilation such a language could be implemented by extending the components for various



phases of compilation and reusing components that don't require changes. Such a facility might also be useful for implementing the multitude of languages related to XML. XML languages usually preserve the basic syntax of XML but add extra semantic requirements. These languages can be implemented quickly and still benefit from the facilities based on the XML language module.

**[0027]** In one embodiment the invention may include tools to speed the development of language modules. These tools may automate the creation of common tasks. In particular automatic generators are common in the art for both parsers and scanners and these tools can make the development of simple languages very rapid. Scanner generators are given a lexical specification, which defines the types of tokens allowed in a particular language and produce code for generating scanners. Likewise, parser generators take a grammar for a programming language and produce a parser that recognizes that grammar. Tools provided with the compiler framework may automatically create components that are compatible with the compiler framework and provide proper interfaces on those components. Tools provided with the compiler framework may also implement robust error correction mechanisms so that the created language modules are suitable for use with all clients.

**[0028]** The compiling system may provide interfaces to provide services and information to various clients. A client may require information about a particular source file or a project. A client may also invoke the compilation of a particular source file or an entire project. A client may also wish to change source files and notify the compiler framework that the source files have changed.

**[0029]** In one embodiment the client may be an integrated development environment (IDE) which allows a developer to work on a project. These facilities may rely on the compiler network to obtain information about the project. The IDE may include facilities for examining the

contents of a project, including browsing the files in a project or browsing the class hierarchy in the project. The IDE may also include an error display for showing the errors in the project.

[0030] In an embodiment that includes an IDE, the IDE may include a source code editor that allows the user to edit source files that are part of the project. The source code editor may wish to request language information about various portions of the source code from the compiling system. This information may be provided by the compiler framework or by the language modules directly.

[0031] A source code editor in an IDE may be adapted to edit source files containing nested languages. The source code editor may request information about the start and end of nested languages from the compiler framework, as well as information about the various different languages in the source file.

[0032] In an interactive embodiment, the compiler framework might provide an interface allowing clients to inform the compiler framework that the files in the project have changed. The compiler framework may subsequently recompile the changed files and any files that depend on them, by obtaining dependency information from the type cache which may be maintained by the compiler framework.

[0033] In another embodiment the client may be a command-line shell. This shell may request that the compiler framework compile a set of files and produce an executable or a library. If the compilation fails, the shell may request a list of errors from the compiler framework so it can display them to the user on the console.

[0034] According to the teachings of the present invention, a software system is created that allows for a compiler that supports both multiple languages and multiple clients. The present system allows for the relatively easy addition of support for new programming languages. Such a system allows for the creation of a flexible development environment that is suitable to the needs of modern programmers who are often working in multiple programming languages and

frequently end up creating new programming languages in order to satisfy the requirements of their current project.

[0035] One embodiment may be implemented using a conventional general purpose or a specialized digital computer or microprocessor(s) programmed according to the teachings of the present disclosure, as will be apparent to those skilled in the computer art. Appropriate software coding can readily be prepared by skilled programmers based on the teachings of the present disclosure, as will be apparent to those skilled in the software art. The invention may also be implemented by the preparation of integrated circuits or by interconnecting an appropriate network of conventional component circuits, as will be readily apparent to those skilled in the art.

[0036] One embodiment includes a computer program product which is a storage medium (media) having instructions stored thereon/in which can be used to program a computer to perform any of the features presented herein. The storage medium can include, but is not limited to, any type of disk including floppy disks, optical discs, DVD, CD-ROMs, micro drive, and magneto-optical disks, ROMs, RAMs, EPROMs, EEPROMs, DRAMs, VRAMs, flash memory devices, magnetic or optical cards, nanosystems (including molecular memory ICs), or any type of media or device suitable for storing instructions and/or data.

[0037] Stored on any one of the computer readable medium (media), the present invention includes software for controlling both the hardware of the general purpose/specialized computer or microprocessor, and for enabling the computer or microprocessor to interact with a human user or other mechanism utilizing the results of the present invention. Such software may include, but is not limited to, device drivers, operating systems, execution environments/containers, and applications.

[0038] The foregoing description of the preferred embodiments of the present invention has been provided for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise forms disclosed. Many modifications and

variations will be apparent to the practitioner skilled in the art. Embodiments were chosen and described in order to best describe the principles of the invention and its practical application, thereby enabling others skilled in the art to understand the invention, the various embodiments and with various modifications that are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the following claims and their equivalents.